

# Work Sample Tests: Their Relationship with Job Performance and Job Experience

## Tests de Muestras de Trabajo: Relación con el Desempeño y la Experiencia del Puesto

Nuno Rodrigues and Teresa Rebelo  
University of Coimbra

**Resumen.** Este artículo examina la relación entre los tests de muestras de trabajo (TMT) y la experiencia laboral (EL) con el desempeño de tarea y el desempeño contextual. En hipótesis, los TMT y la EL están relacionados con el desempeño de tarea, porque ambas medidas están relacionadas con las tareas y otros elementos técnicos intrínsecos del trabajo. Sin embargo, esperamos una relación no significativa entre el desempeño contextual y los TMT y la EL. Usando una muestra de 60 trabajadores de ensamblaje, los resultados sugieren que las muestras de trabajo son predictores válidos del desempeño de tarea, pero que no predicen el desempeño contextual. Con respecto a la experiencia laboral, los resultados indican una correlación moderada con las dos dimensiones de desempeño. Además, se encontró una correlación elevada entre los dos predictores. Finalmente, se discuten las implicaciones de estos resultados en relación con los dos predictores.

*Palabras clave:* validez predictiva, tests de muestras de trabajo, experiencia laboral, desempeño de tarea, desempeño contextual.

**Abstract.** This article examines the relationship between a work sample test (WST) and a measure of job experience (JE) with task and contextual performance. Hypothetically, WST and JE are related to task performance, because both are connected with working on tasks and other intrinsic technical elements of the job. Nevertheless, a non-significant relationship between contextual performance with WST and JE is to be expected. Using a sample of 60 assembly workers, the results suggest that work samples are indeed valid predictors of task performance but they do not predict contextual performance. With respect to job experience, the results reveal a moderate correlation with both dimensions of performance. Furthermore, a large correlation between both predictors was found. Implications of these results concerning the two predictors under study are discussed.

*Key words:* predictive validity, work sample tests, job experience, task performance, contextual performance.

Advances in research of personnel selection occurring over the last hundred years, based on both primary studies and meta-analytic evidence, have demonstrated that organizations will greatly benefit from the use of validated selection methods (Hausknecht, Day, & Thomas, 2004; Schmidt & Hunter, 1998; Van Iddekinge & Ployhart, 2008;). Indeed, selection systems should include valid selection tools and procedures, due to their important contribution to the achievement of higher standards of performance at an individual, group and organizational level (Cook, 2004; Schmidt & Hunter, 1998; Van Iddekinge & Ployhart, 2008). Furthermore, the legal defensibility of selection decisions relies on the inherent predictive

validity of the methods used to support the decision-making process (Van Iddekinge & Ployhart, 2008).

The predictive validity of a selection method is related to its ability to predict relevant criteria, such as the most often studied job performance and training performance, or additional ones, like wages, accidents, turnover and professional status change (Cook, 2004; Schmidt & Hunter, 1998). The prediction of job performance, particularly, has a great relevance in the determination of the success of selection decisions, because effective job performance is related to desirable organizational outcomes (i.e., larger outputs of a product or a service) (Avis, Kudisch, & Fortunato, 2002; Robertson & Smith, 2001; Viswesvaran, 2001).

In the scope of predictive validity research in selection, this study examines the relationships between work sample tests and job experience as predictors and job performance as a criterion, as well as the relationship between the two predictors. The job performance

---

The research reported in this paper was supported by a Portuguese Foundation for Science and Technology (FCT) Doctoral Grant to the first author. Correspondence should be sent to Nuno Rodrigues, Faculty of Psychology and Sciences of Education, Rua do Colégio Novo, Apartado 6153, 3001-802 Coimbra, Portugal. E-mail: nuno.rodrigues002@gmail.com

criterion is operationalized through the task and contextual performance dimensions proposed by Borman and Motowidlo (1993).

### ***The expanded job performance criterion: Task performance and Contextual performance***

Nowadays, due to the relevant empirical and theoretical work published in the literature on selection, there is general recognition that the performance construct has a complex and inherently multidimensional nature (Austin, 1964; Campbell, 1990; Murphy & Shiarella, 1997; Viswesvaran, 2001). On the other hand, some authors claim that different individual variables are correlated with specific performance dimensions and particular aspects (or facets) (Borman & Motowidlo, 1993; Borman & Motowidlo, 1997; Van Scotter & Motowidlo, 1996; Viswesvaran, 2001). In fact, some efforts have been made to carry out current research on the individual job performance domain, based on the general recognition that this non-unitary criterion needs further conceptual and empirical clarification in order to create additional lines of sight on its relationships with several predictor variables (Robertson & Smith, 2001; Viswesvaran, 2001). An important path to a better understanding of individual job performance and its latent, complex and multidimensional nature is concerned with the study of the dimensions that make up the content of this construct (Austin, 1964; Borman, Hanson, & Hedge, 1997; Campbell, 1990; Murphy & Shiarella, 1997; Viswesvaran, 2001). Having identified its underlying dimensions through the examination of the individual performance manifestations, a further step should be made in order to understand which performance dimensions could be generalized across work settings and those which can vary in different workplaces (Borman et al., 1997; Viswesvaran, 2001).

Indeed, despite some empirical and conceptual work carried out until the present, we are yet to find a convincing degree of consensus among authors about which dimensions of performance should be chosen to represent performance for most jobs (Borman et al., 1997; Borman & Motowidlo, 1993; Hatrup, O'Connell, & Wingate, 1999). However, the distinction proposed by Borman and Motowidlo (1993) between task and contextual performance represents one attempt to clarify the dimensions that could make up the broad construct of job performance. According to the above mentioned distinction, task performance can be defined as a behavior that serves and maintains the execution of the role's pre-described activities, contributing to the efficiency of the technical core of the organization's functioning, either directly by direct implementation of a technological process, or indirectly by providing materials or services (Robertson & Smith, 2001). On the other hand, contextual performance is related to the concepts of organizational citizenship, extra-role

behavior and pro-social organizational behavior and is defined as behavior that maintains or improves the social and organizational context of the task core. Some previous research has shown that different individual variables assessed by the selection methods are better predictors of specific performance dimensions, and even of particular elements or facets (Borman & Motowidlo, 1993; 1997; Van Scotter & Motowidlo, 1996; Viswesvaran, 2001). In accordance with Borman and Motowidlo (1997), contextual performance will be best predicted by personality, motivation and personal orientation differences, whereas other variables, like abilities, knowledge, and experience will be better predictors of task performance.

As we already noted, one way of refining research is to include dimensions of performance, instead of one unitary measure of performance, as criteria in the context of selection studies, because it contributes to a better understanding of the impact of different selection methods on different elements of performance criteria. Due to the fact that our research is focused on production assembly workers who carry out their specific tasks in the assembly line individually, but who at the same time are integrated into work teams, the two facets of performance above described are crucial types of performance in the context under study. For this reason, we have chosen them as performance criteria.

In spite of the conceptual clarification that results from the separation between the two described performance dimensions, there is little field research supporting the empirical independence of task and contextual performance (Bott, Svyantek, Goodman, & Bernal, 2003). Thus, an additional objective of our study is to test the theoretical rationale provided by Borman and Motowidlo (1993), through examination of work sample tests and job experience as stronger antecedent variables of task performance, compared to contextual performance. These predictors will be discussed in the following sections along with their respective hypotheses.

### ***Defining Work Sample Tests***

Prior research has demonstrated that work sample tests constitute a selection method that should be placed in the group of the most valid job performance predictors (Hunter & Hunter, 1984; Roth, Bobko, & McFarland, 2005; Schmidt & Hunter, 1998). The validity of work samples is recognized both by researchers and the professionals involved in selection activities, such as human resources practitioners and managers (Terpstra, Kethley, & Foley, 2000). According to Ployhart, Schneider, and Smith (2006, p. 538) "*a work sample test is a test in which the applicant performs a selected set of actual tasks that are physically and/or psychologically similar to those performed on the job*".

The definition cited above was also adopted by Roth, Bobko, McFarland and Buster (2008) who highlighted, in addition, that the structure of the work sample scoring systems and procedures should be elaborated under the orientation of experts in the specific job to which the test corresponds (Schneider & Schmitt, 1986). It should be also noted that work sample testing represents “*an approach or rationale for the assessment of individuals’ current or likely future job performance*” (Callinan & Robertson, 2000, p. 248) rather than a unitary method. Therefore, the work sample is necessarily specific to a particular job and represents a hands-on performance test that encompasses an evaluation situation that replicates the real job conditions, in which the candidate has to perform a set of job-related tasks or solve similar job problems (Cook, 2004; Roth et al., 2008). The job specificity of work sample tests is also emphasized by Salgado, Ones and Viswesvaran (2001, p. 178) who state that “*simulations can be more sophisticated in capturing the psychological and physical aspects of the work settings. We consider work sample tests in this section*”. Moreover, this high level of fit between jobs tasks and/or problem solving activities represents an idiosyncratic characteristic of work sample tests (Guion, 1998).

Another relevant point concerning work samples is related to the distinction that must be made between these tests and other, different sorts of tests, namely “performance tests”, in which applicants perform the job itself during a specific time. Internships and probationary periods are examples of this last kind of test and therefore cannot be classified as work samples (Heneman & Judge, 2003; Roth, Bobko, & McFarland, 2005). Finally, it is also important to notice that, despite the frequent use of work sample tests as a predictor measure, they can be also used as a criterion measure as well, to validate important organizational outcomes, like training success (Callinan & Robertson, 2000).

### **Validity of Work Sample Tests**

One of the first meta-analyses concerning work sample validity was conducted by Hunter & Hunter (1984). This seminal study led us to conclude that for workers who already know the job, the validity of work sample tests was .54 on job performance (corrected for criterion unreliability, but not corrected for range restriction). This estimate is larger than the estimate validity found for the best individual performance predictor, general mental ability ( $r=.51$  corrected for range restriction and measurement error in criterion). A previous “smaller scale” meta-analysis study was conducted by Hunter (1983, cited by Roth et al., 2005) in a sample of a non-military studies, and it established a mean correlation of .42 ( $K=7$ ,  $N=1,790$ ) between work samples and supervisory ratings, after correction for criterion unreliability. An additional esti-

mate was also calculated by the author using military studies ( $r=.27$ ,  $K=4$ ,  $N=1474$ ).

Schmitt, Gooding, Noe, and Kirsh (1984) conducted another relevant study using primary studies published between 1964 and 1984 in the Journal of Applied Psychology and the Journal of Personnel Psychology and found an uncorrected validity of .32 ( $K=7$ ,  $N=382$ ) when the criteria was job performance ratings. Some additional non-corrected coefficients were calculated for achievement/grades ( $r=.31$ ,  $K=3$ ,  $n=95$ ) and wages ( $r=.44$ ,  $K=4$ ,  $n=1191$ ) criteria (Salgado et al., 2001). A decade later, Russel and Dean (1994) extended the scope of Schmitt et al.’s (1984) previous study with subsequent published research, and the observed validity estimate was .37 ( $K=20$ ,  $N=3894$ ) for validity of work samples in job performance in a variety of jobs. More recently, the magnitude of the validity coefficients (corrected for criterion unreliability but not for range restriction) obtained in a study based on meta-analytic procedures conducted by Roth et al.’s (2005) was substantially lower ( $r=.33$ ) than the magnitude of the estimates resulting from the initial work of Hunter and Hunter (1984) ( $r=.54$ ). Based on these results, Roth et al. (2005) emphasized that work sample validity could be noticeably lower than previously thought and therefore some organizations may be overestimating the validity properties of work samples. Despite the lower validity estimates that were obtained in more recent studies, which have overcome important limitations related with conceptual and methodological problems (Asher & Sciarrino, 1974) on the one hand, and the limited scope on the other, in general, the directionality of the results of the studies converge on an indication that work sample tests are valid predictors of job performance.

On the basis of this information, we expect that the work sample tests will be related to performance. However, assuming that the work samples specifically created for this research mirror critical technical job operations (e.g. welding, cutting, painting) which are performed individually and do not include any interpersonal/contextual component, we hypothesized that:

*H1: Work sample ratings will be positive and significantly related to task performance.*

*H2: Work sample ratings will not be significantly related to contextual performance.*

### **Other advantages and some disadvantages of work sample tests beyond their validity**

Work sample tests have been widely used as a selection method since the beginning of the 20<sup>th</sup> Century (Callinan & Robertson, 2000; Salgado, Ones, & Viswesvaran, 2001). A number of additional favorable arguments that go beyond their predictive validity justify their frequent utilization for selection purposes. One of these positive attributes is related to the incre-

mental validity of work samples. Schmidt and Hunter (1998) focused on this specific issue and reported that supplementing a general mental ability test with a work sample was responsible for a considerable increase of .12 in validity. From the point of view of Salgado et al (2001, p. 180) "*this is probably one of the largest amounts of incremental validity shown by a personnel selection method*". The positive applicant reactions (Hausknecht, Day, & Thomas, 2004) and the resulting lower levels of adverse impact on ethnic groups, compared to paper-and-pencil tests of cognitive ability, constitute other remarkable positive characteristics of this method (Callinan & Robertson, 2000; Salgado, Ones, & Viswesvaran, 2001). With respect to the low level of adverse impact of work samples, it should be noted that this is related to the specific constructs that can be assessed by the work sample (Hough, Oswald, & Ployhart, 2001). In fact, an increase in the degree of adverse impact should be expected if the cognitive load of the exercises that are included in the simulation/work sample increases (Goldstein, Yusko, Braverman, Smith, & Chung, 1998; Roth et al., 2008).

Although the positive characteristics associated with work samples can generate great enthusiasm, it is important to summarize other aspects that preclude or, at least, constrain, their use for selection purposes. One of these aspects is related to the fact that these tests can only be used with experienced applicants (Schmidt & Hunter, 1998). Another criticism involves the argument that work samples are suitable for evaluation of concrete skills, but they do not work so well if the job includes an abstract and diverse task component or involves interaction with other people (Cook, 2004). On the other hand, due to its job specificity, the corresponding development and application process is undoubtedly costly and complex (Callinan & Robertson, 2000). Moreover, the necessary individual administration and the need for experts to rate the applicant's performance in the test make work samples procedures both costly and time consuming (Callinan & Robertson, 2000; Cook, 2004). Thus, logistical, financial, and safety issues play an important role in the viability of work sample use in personnel selection. Some authors go further in their criticism of work samples, arguing that although the predictive validity of work samples is high at the moment of the selection process, it seems to weaken over time to a greater extent than other selection measures (Robertson & Kandola, 1982; Callinan & Robertson, 2000). Thus, job learning and adaptation for successful performance over time may be more related to general underlying abilities (Callinan & Robertson, 2000).

### ***Relationships of work samples with other constructs: what do they really measure?***

The intrinsic multifaceted nature of work samples gives them the possibility of capturing more criterion

variance than other predictor measures (Schmidt, Clause, & Pulakos, 1996). However, the research questions concerning the underlying constructs measured by a work sample remain unsatisfactorily answered (Salgado et al., 2001). Actually, due to the fact that work sample tests are developed in order to replicate the nature of the job, instead of focusing on a specific and well-defined construct, investigation of the constructs measured on a work sample constitutes a complex matter (Roth et al., 2008).

Schmidt and Hunter (1992) presented a process model stating that work sample performance is explained by the relationships between motivation, job experience and cognitive ability. The motivation component is somewhat related to the face validity of these tests, which contributes to the applicant's perception of the test as a fair opportunity to show his abilities (Smith, 1991). Schmidt and Hunter (1998) made a contribution to this topic by reporting a correlation of .38 between general mental ability and work samples. Roth et al. (2005) indicated an observed correlation of .32 ( $K=43$ ,  $N=17,563$ ) which rose to .38 when the estimation was corrected for work sample unreliability and .40 when both types of tests were corrected for unreliability. Schmidt and Hunter's model suggests that cognitive ability impacts on work sample performance through its effects on the acquisition of job knowledge, because cognitive ability increases the speed of acquiring job knowledge processes (Schmidt, Hunter, & Outerbridge, 1986). According to this point of view, the increase in work sample performance is due to the previous acquisition of job knowledge (Campbell, Casser, & Oswald, 1996). This model has received strong support mainly in military studies (Borman, White, Pulakos, & Oppler, 1991; McCloy, Campbell, & Cudeck, 1994), but other studies with civilian organizations have also supported it (Hunter, 1983). More recently, Roth et al. (2008) reported that work samples are saturated on several constructs such as cognitive ability, job knowledge, and social skills, whereas personality, interests, and organizational fit variables are not related to this selection method.

### ***Job Experience, work sample tests and job performance***

Job experience is another variable that appeared to be correlated with work samples and job performance in several studies (Kolz, MacFarland, & Silverman, 1998; Quinones, Ford, & Teachout, 1995; Schmidt & Hunter, 1998; Schmidt, Hunter, & Outerbridge, 1986). Job experience was usually defined in quantitative terms as the number of years in the same or similar job (Quinones et al., 1995; Schmidt & Hunter, 1998). Previous research suggests, however, that the number of times a person performs a job task is more highly

related to job performance than the number of years in the job (Quinones et al., 1995). Prior research also supports that the major impact of work experience both on work sample performance and job performance may be established through its direct effect on the acquisition of job knowledge (Quinones et al., 1999; Schmidt, Hunter, & Outerbridge, 1986). Quinones et al. (1995) found a correlation of .39 (corrected for sampling error and criterion unreliability) between measures of hard work samples and job experience. Based on this information and on the Schmidt and Hunter (1992) model, which postulates that work sample performance depends on job experience and other relevant variables of the applicant (motivation and cognitive ability), we formulate the following hypothesis:

*H3: Job experience will be positively related to work sample ratings.*

Regarding job performance, and according to Bott et al. (2003), job experience impacts on task and contextual performance in distinct ways. Based on the assumption that task performance reflects proficiency in carrying out tasks detailed in a formal job description, it will increase as employees obtain specific job knowledge that allows them to perform the tasks at a higher level (Hatrup et al., 1998). Conversely, contextual performance includes behaviors like helping work colleagues with a heavy workload, cooperating, following rules with enthusiasm, or volunteering for non-formal duties (Borman & Motowidlo, 1997; Hatrup et al., 1998). The learning of these behavioral conducts is also developed through various experiences that occur before commencing professional activity and in activities outside the job, being then transferred to the individual's job experiences (Bott et al., 2003). Hence, differences in contextual performance are more likely to be correlated with differences in personality and interpersonal orientation than with job experiences. On the other hand, the jobs analyzed in the present research are highly structured, basically involving a set of routine activities, and, in addition, all the participants had had experience in similar tasks in previous jobs. Thus, in this particular case, the measure of job experience used (number of years in the job) is probably more highly related to the number of times each task was performed than to aspects regarding contextual performance. Relying on this rationale we hypothesize that:

*H4: Job experience will be positively and significantly related to task performance.*

*H5: Job experience will not be significantly related to contextual performance.*

## Method

### *Participants and setting*

The data were collected in the production department of a Portuguese bus bodywork assembly compa-

ny, located in the central region of the country. At the time of data collection (2007), this department consisted of approximately 200 employees, distributed among work teams. In fact, the assembly workers in this production department carry out the majority of their particular formal tasks individually on the assembly line, due to the specific and technical nature of activities (welding, cutting, painting). However, every functional production section has different workers, forming a work team. The members of each team work towards the same goals concerning bus bodywork production and with the same supervisor. In spite of the individual work on the assembly line, the members of each team have periodical meetings for discussing work issues, where they are encouraged to participate and make suggestions. The sample of this study is made up of 60 of these assembly workers, belonging to four different sections: paint, sheet metal, bus structure assembly, and fittings and finishing. Besides their different positions in the bus production line, all the individuals are operators, with no leadership responsibilities and performing operative tasks. Of the 60 participants, 23 (38%) were hired applicants and 37 (62%) were already employees of the company or incumbents. The majority of participants were male (93%), under 40 years old (82%) and with nine years of schooling or less (83%). In terms of job experience, 10% of the workers had one year or less, 38% from 2 to 5 years, 23% from 6 to 10 years, 10% from 11 to 15 years and the rest (19%) had 16 years or more.

### *Data collection procedures*

The data was collected in two phases. The first phase took place at the end of 2006 and the beginning of 2007, and involved the administration of work sample tests. The 23 applicants did the work sample tests during the respective selection process. The data collection of the predictors relating to the incumbents occurred shortly after the data collection from applicants. The incumbents did the work samples tests inside their working hours. In order to ensure there was sufficient opportunity for the supervisors to observe the applicants' performance, the second phase took place in July 2007, where each participant's direct supervisor assessed his job performance.

### *Measures*

#### *Work sample tests.*

The work sample tests used in this study were specially conceived by the research team, with the collaboration of supervisors and experienced workers in each production section. Six work samples were created, corresponding to the six principal jobs requested by the

production sections to which the participants belonged (welder in bus structure assembly section; carpenter, metal worker and panel beater in sheet metal section; painter in paint section; and fittings and finishing tasks in their respective section). As we can see, all the work samples referred to technical tasks related to bus assembly. Each participant did only one work sample test, according to his current job in the production department or, in the case of applicants, related to the job they had applied for in the selection process.

All the work samples were structured at three levels of difficulty, each one comprising tasks that implied a certain level of technical proficiency, regarding the job under evaluation. More difficult levels corresponded to more sophisticated technical tasks, where satisfactory performance required higher levels of proficiency. The procedures of each work sample were standardized, through an administration protocol created by the researchers and supervisors. In consequence, for the same work samples, the same instructions were given to participants and they were carried out in the same conditions, concerning tools, security, equipment and time. In order to maintain standardization and also facilitate the evaluation process, assessment grids were drawn up for each type of work sample. At each level of difficulty, a set of behaviors related to the task under evaluation was described and the supervisor rated them on a seven-point scale ranging from 1- *very poor* to 7 - *excellent*. Based on these grids, the final mark of each participant was calculated, first, by the mean of the item rates at each level of difficulty (mark obtained in each level), and, afterwards, by the mean of the marks achieved at each level of difficulty (when, for instance, a participant did not pass to the last level, the mark for this level was 0). To enhance the reliability of these tests, supervisors were trained to carry out and rate them, the work samples and the assessment grids were previously tested with a small group of workers not belonging to our sample, and one of the researchers was present for administration of the work samples to participants in the study.

#### *Job experience.*

Concerning applicants, information about job experience was obtained from the application form and checked in the selection interview. Information about the job experience of company incumbents was provided by the human resources department. Job experience was measured in a quantitative manner, corresponding to the time (years and months) of the applicant's job experience, and not by the number of times to each task was performed, which, at a first glance, could add other richness to the quantitative measure. This kind of measure (years and months of job experience) was taken because, in the present study, as we have already mentioned, jobs under evaluation are

highly structured, including a limited number of routine activities and all subjects had experience in the job. Regarding these types of jobs, the number of years in the job is probably highly related to the number of times each task is performed. Thus, the number of years/months would represent a more readily available, practical, and less intrusive measure of experience than counting the number of times a task is performed (Kolz et al., 1998).

#### *Job performance.*

To measure job performance, we adapted two scales according to the two dimensions of job performance proposed by Borman and Motowidlo (1993): contextual and task performance. The "translate - translate back" method was used to adapt both scales to Portuguese. The scales were included in a single questionnaire, where supervisors rated the individual job performance of the participants. Ratings of task performance were obtained from supervisors using the nine items, a seven-point Likert scale originally developed by Bott, Svyantek, Goodman and Bernal (2003), with higher ratings indicating better performance. These items are centred on issues like fulfilment of objectives required by the job, competence to do all tasks, and the potential for promotion. "Achieves the objectives of the job" and "demonstrates expertise in all job-related tasks" are examples of items of this measure. We chose to adapt this scale because the items, as a whole, are very clear and fit the work context of the participants. Besides, the exploratory factor analysis carried out by the authors of the scale indicates a single factor, with the nine items together and an  $\alpha = .93$ . Supervisors rated operators' contextual performance on a seven-point Likert scale with nine items, (1- *Strongly Disagree*; 7- *Strongly Agree*), with higher ratings indicating better performance. These nine items were used by Morgeson, Reider and Campion (2005), which were taken from Moorman and Blakely (1995), Motowidlo and Van Scotter (1994), and Van Scotter and Motowidlo (1996). Morgenson et al. (2005) introduced minor modifications to the items to make them to refer explicitly to a team. Since the participants in this study were organized in work teams and the scale had also been used previously in the context of a steel corporation, these aspects, together with the clarity of items, led us to choose this measure. Items include issues such as cooperating with team members, going out of his or her way to help other team members, and are related to interpersonal facilitation, interpersonal helping, job dedication, and individual initiative (Morgeson et al., 2005). Examples of items are "cooperates with others in the team" and "offers to help other team members accomplish their work". An exploratory factor analysis carried out by these authors indicates that a single fac-

tor accounts for the nine items with an internal consistency reliability coefficient of .98.

Assuming some consensus around the idea that both task and contextual performance contribute to overall job performance, and they are related to each other, it is possible to integrate them in one general measure (Murphy & Shiarella, 1997). Thus, although there were no formal hypotheses associated with overall performance, we have decided to take into account in this study a measure of overall performance as an additional criterion in order to provide some extra information, complementary to the analyses carried out considering contextual and task performance. As a composite measure, the overall job performance score corresponds to the sum of the task and contextual performance scores obtained by each participant.

### *Data analysis: previous procedures*

Factor and reliability analyses of the job performance measures were carried out with a sample of a hundred subjects ( $N=100$ ). In fact, these procedures implied the enhancement of the initial sample size of 60 participants with additional data collection. To do so, we asked supervisors to assess 40 more workers with the same job performance questionnaire. This additional data served only to allow the feasibility of these analyses. Exploratory factor analyses considering the 18 items as a whole (nine from task performance and nine from contextual performance) revealed a bi-dimensional structure, as expected. However, we had to withdraw three items pertaining to the original scale of task performance, due to their high and equivalent loading in both factors (they loaded above .50 on the two factors, but their loadings differed less than .10). Without these three items, the factorial structure explains 77.9% of the total variance and all the items have loadings above .70 and communalities above .50. The nine original items of contextual performance remained together in one factor, explaining 33.6% of variance with an  $\alpha = .97$ . The other factor is made up of the remaining six items of task performance, accounting for 44.3% of explained variance and with an  $\alpha = .94$ . Task and contextual performance appeared inter-correlated ( $r=.73$ ,  $N=100$ ) and the internal consistency coefficient concerning all the 15 items together (overall performance) was .97.

In order to analyze if our job performance measure was suitable for the participants' work context, we correlated the overall performance measure with the results of the performance appraisal process carried out by the company regarding the year of data collection. The correlation coefficient between the two measures ( $r = .66$ ,  $p < .001$ ) showed a good level of convergence between them.

As already stated, due to external constraints to our research, we had to collect data not only from appli-

cants, but also from a set of incumbents of the company, in order to enlarge the sample and allow the research to proceed. However, we were aware that these two groups of participants could be different in the variables under study, and consequently, that these differences could, to some extent, produce biased results. Due to this, we checked for significant differences between the two groups of participants in all the variables that we were going to study. The  $t$  tests carried out showed there were no significant differences between hired applicants and incumbents concerning work sample ratings ( $t(58) = -1,982$ ,  $p = .052$ ), job experience ( $t(58) = -1,874$ ,  $p = .071$ ), task performance ( $t(58) = -1,976$ ,  $p = .053$ ), contextual performance ( $t(58) = -0,437$ ,  $p = .665$ ), and overall performance ( $t(58) = -0,1217$ ,  $p = .232$ ). This procedure was carried out to increase the reliability level in treating both groups as a single one, i.e. as a whole and homogeneous sample in data analysis. Besides the effort to enlarge the sample to 60 participants, statistical power to identify a significant correlation is 34% to detect a moderate small effect ( $r = .20$ ,  $p < .05$ , two-tailed) and 67% to detect a medium-size effect ( $r = .30$ ,  $p < .05$ , two-tailed). Finally, results of the evaluation of statistical assumptions led to transforming the job experience variable using a logarithmic transformation, due to its positive skewness (Tabachnick & Fidell, 2001).

### **Results**

Table 1 presents means, standard deviations, and correlations between the variables under study. As can be seen, work samples are positively and significantly related to task performance and to overall performance, but not to contextual performance. In addition, the highest magnitude is with task performance ( $r = .45$ ), against a moderate small size coefficient with contextual performance ( $r = .23$ ). Apart from the weak statistical power of the present sample to detect significant moderate small effect sizes, the differences between these coefficient sizes showed that the relationship between work samples is stronger with the task performance facet than the contextual one. Thus, the results support our hypotheses 1 and 2, which stated that the substantive relationship between work samples and the two performance dimensions under study would be with task performance.

Regarding the relationship between job experience and work sample ratings, as hypothesized in H3, both constructs are inter-correlated, and with a large coefficient ( $r = .53$ ), giving support to the idea that work samples could indeed be influenced by job experience. Concerning years of job experience and its relationship with job performance, this variable is not significantly related to any dimension of performance (task and contextual) nor with overall performance, presenting moderate small size correlations ( $r = .22$ ,  $r = .23$ , and  $r = .24$ ,

respectively). Therefore, these results support hypothesis 5 (job experience is not significantly related to contextual performance), but do not support hypothesis 4, since job experience is not significantly related to task performance either. Regarding task performance, years of job experience is correlated to this dimension with a lower magnitude, in comparison to work sample tests ( $r=.22$  against  $r=.45$ , respectively).

tious, predominantly related to the task performance dimension. The difference of magnitude of the relationships found mostly corresponds to our expectations, since the work sample tests that were specially developed in the scope of this research have a high technical and proficiency task core and do not include any component that focuses on evaluation of personality/disposition or interpersonal variables.

Table 1. Means, Standard Deviations, and Correlations among Study Variables

Variable	M	SD	1	2	3	4	5
1. Work Sample Test	4.87	1.09	—	—	—	—	—
2. Log. Job Experience	0.71	0.55	.53**	—	—	—	—
3. Task Performance	4.56	0.64	.45**	.22	—	—	—
4. Contextual Performance	4.13	0.59	.23	.23	.59**	—	—
5. Overall Performance	4.26	0.55	.35**	.24	.84**	.91**	—

Note. N = 60; \*\* $p < .01$ ; \*  $p < .05$  (two-tailed).

## Discussion

The aim of this study was twofold. We intended to study the predictive contribution of work samples and job experience to job performance by examining the pattern of relationships between these predictors and the dimensions of task and contextual performance, and also explore the relationship between these two predictors.

In contrast to what has happened in the past, the study of performance recurring to a non-unitary conceptualization is increasing in importance, because of the important contributions that can be provided to the study of the conceptual structure of performance and also of the variables that can impact on it by the assessment of the links between different sources of individual variability, most of them assessed through well-established predictors (Robertson & Smith, 2001; Salgado, Ones, & Viswesvaran, 2001). Taking into consideration what we emphasized above, we adopted a multidimensional operationalization of performance focused on the dimensions of task and contextual performance proposed by Borman and Motowidlo (1993) to provide a better understanding of the impact of work samples and job experience on the prediction of job performance.

After these preliminary notes, we will proceed to the discussion of the results of this study and, at the same time, will also list some of its limitations and add some recommendations for future research when appropriate. As described above, our results concerning work samples are consistent with our hypotheses. In fact, this measure is related to the task performance dimension but not to the contextual. These results seem to suggest that the constructs which impact on work sample performance are only related, or to be more cau-

As previous research provided evidence of relationships between work samples, job experience and job knowledge, but not connections with other constructs such as personality variables, interests and organizational fit, it could be plausible to state that work samples have greater potential to predict task dimension rather than contextual performance, because they often targeted constructs related to task performance (Roth et al., 2008; Schmidt & Hunter, 1992; Roth et al., 2005; Hunter, 1983; Ones & Viswesvaran, 1998). From our point of view, future research especially related to work samples should address the limitations concerning the lack of comprehensive studies in the personnel selection literature about its construct validity. This kind of research is crucial for a better understanding of the relationships between work samples and criteria, and at the same time it will be easier to find which constructs relies on the common core of the different work sample tests.

Concerning contextual performance, prior research gradually gave most of the merit of the prediction of this dimension to individual voluntary predispositions, and consequently, to personality variables (Borman, Penner, Allen, & Motowidlo, 2001; Borman & Motowidlo, 1997; Borman & Motowidlo, 1993). Therefore, we did not expect any significant relationship between this specific predictor (work samples) and contextual performance, and our results support our hypothesis. So, at this point, we believe it should be emphasized that one of the most desirable characteristics of the use of work samples for selection purposes – its potential for predictive validity – was demonstrated once again in our research, but concerning one dimension of performance: task performance. These results gain importance inasmuch as, despite the theoretical merits of the emergence of contextual per-

formance as a dimension of performance (homologous to task performance), empirical support for this conceptual distinction has been inconclusive. Few field studies provide empirical support for independence between the dimensions of task and contextual performance (Motowidlo & Van Scotter, 1994; Van Scotter & Motowidlo, 1996; Bott et al., 2003). Actually, in our sample these two dimensions are inter-correlated, but even so, the magnitude of work samples with task and contextual performance differ one from the other. So, they suggest that these two dimensions are, indeed, inter-related, but cover different and specific aspects of job performance.

As additional information, we decided to include a measure of overall performance. Concerning this supplementary measure, the coefficient obtained for work sample validity was approximately  $r=.35$ . It is interesting to notice that this coefficient is very similar in magnitude to the estimate of Roth et al. (2005) ( $r=.33$ , when measures of job performance were corrected for attenuation). What is more, it is not surprising that this coefficient is lower than the coefficient regarding task performance, but greater than the coefficient with contextual performance, since it is a composite measure, based on the sum of the scores of both dimensions.

Continuing in the scope of the discussion of work sample results, in the present research they were measured with supervisor ratings, and only one rater (one supervisor) in each production sector was available to evaluate both work sample performance and job performance. In these circumstances, the rater could fail to distinguish between multiple dimensions of the performance criteria and the results obtained may be inflated due to halo or common method variance (Van Iddekinge & Ployhart, 2008). We consider this as one limitation of our study so we subscribe to the recommendations of some authors who argue that future research should gather data from different sources (multiple supervisors, peers), and when possible, objective measures to evaluate performance (Bott et al., 2003).

About the relationship between the two predictors under study, as we have already stressed, some authors have emphasized the need to carry out additional research in order to achieve a better understanding of the constructs that are measured in a work sample (Salgado et al., 2001). The work of Schmidt & Hunter (1992), already mentioned, represents one of the few contributions to this topic. According to Schmidt and Hunter's model, the major impact on task proficiency is due to job knowledge, which means that work sample performance is improved by job knowledge. The acquisition of job knowledge is, in turn, influenced by cognitive ability and increased through job experience (Campbell, Casser, & Oswald, 1996). Based on their contributions, one of the goals of this article is to examine the relationship between work samples and job experience.

As we hypothesized, the results show a significant relationship between work sample tests and job experience ( $r=.53$ ). This positive and larger correlation fits Schmidt and Hunter's model, because, as stated above, it suggests that the increase in job performance (some of its aspects are measured through work samples) is probably due to increments in job knowledge acquired during job experience, which contributes directly to better levels of performance. Additionally, this result leads us to emphasize, once again, that further research focusing on what is really measured by work samples is an interesting and necessary way to better understand this important measure in the context of selection.

As for the results related to the other predictor under study – job experience – and its relationship with job performance, we expected that job experience would be significantly, or at least moderately correlated with task performance, but not with contextual performance, based on the assumption that acquired job knowledge contributes in a minor or insignificant way to individual differences in contextual performance, which are frequently related to personal discretion and dispositions that form the worker's personality (Speier & Frese, 1997). Conversely, the correlations obtained between job experience and task and contextual performance, despite their non-statistical significance due to the weak statistical power of our sample size, had similar moderate small magnitudes ( $r=.22$  and  $r=.23$ , for task and contextual dimensions, respectively). Therefore, job experience, in comparison to work samples, seems to be a weaker predictor of task performance.

Concerning job experience and its relationship with job performance, Schmidt and Hunter (1998) point out that the increase in job knowledge, caused by time and types of work experience in the job, mostly occurs during the first five years, approximately. After this period of job knowledge acquisition, the enhancement of job experience produces small or negligible increases in job performance. We checked the type of relationship between these variables in our sample and it does not reveal this pattern, but unfortunately we do not have a sufficiently large sample size to explore this aspect consistently. However, due to the relevance of this matter, future research centered on these variables should pay attention to the possibility of having some kind of non-linear relationship between them.

It is also important to highlight that the quantitative measure of job experience used could be another limitation of present study, because it does not cover the qualitative components of this variable. Therefore, the measure taken in the present study does not cover the multidimensional nature of the job experience construct, thereby probably limiting its predictive power, when we also take into consideration a multidimensional criterion of performance (Quinones, Ford, & Teachout, 1995; Tesluk & Jacobs,

1998). Thus, a more complete measure of job experience should include issues such as opportunities for training in new skills or updating those already acquired, supporting and leading others, challenging tasks previously encountered, and types of problems solved. This sort of job activities shapes job experience and influences how the experience translates into job knowledge, skills and motivation, and consequently affects the employee's performance. Therefore, the use of additional measures of qualitative job experience components undoubtedly constitutes a stimulating orientation for future research because the use of restricted quantitative job experience measures does not represent the entire potential for explanation and prediction of this construct for general, multidimensional performance and other relevant criteria (Tesluk & Jacobs, 1998).

Finally, it will also be interesting in terms of further investigation to replicate this study in similar jobs and also with other sorts of jobs, but with a larger sample in order to contrast the results because, as we have already recognized, we worked with a small sample size, forcing us to interpret our results with caution, in spite of their interest.

In summary, despite the limitations inherent to the present research mentioned above, our results point to the strong relationship between time of job experience and work samples, which, in turn, have revealed themselves to be a good predictor of task performance. The option to treat performance criterion not as one-dimensional measure, but in terms of two of its sub-dimensions, gave us the opportunity to show that work samples are much more related to one of them (task performance) and less related with the other (contextual). Hence, when we consider as criteria various dimensions of performance, instead of a unique, overall measure of performance, it is certainly a means to refine research focused on the predictive validity of selection methods.

## References

- Asher, J. J., & Sciarino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology*, *27*, 519-533.
- Astin, A. (1964). Criterion-centered research. *Educational and Psychological Measurement*, *24*, 807-822.
- Avis, J. M., Kudisch, J. D., & Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology*, *17*, 87-105.
- Bobko, P., Roth, P.L., & Buster, M. (2005). Work sample selection tests and expected reduction in adverse impact: A cautionary note. *International Journal of Selection and Assessment*, *13*, 1-10.
- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmidt, W. C. Borman, A. Howard, A. Kraut, D. Ilgen, B. Schneider, & S. Zedeck (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass Publishers.
- Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, *10*(2), 99-109.
- Borman, W. C., Hanson, M. A., & Hedge, J. W. (1997). Personnel Selection. *Annual Reviews of Psychology*, *48*, 299-337.
- Borman, W.C., Penner, L.A., Allen, T.D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, *9*, 52-69.
- Borman, W.C., White, L.A., Pulakos, E.D., & Oppler, S. H. (1991). Models of supervisor job performance ratings. *Journal of Applied Psychology*, *76*, 863-872.
- Bott, J.P., Svyantek, D. J., Goodman, S.A., & Bernal, D.S. (2003). Expanding the performance domain: Who says nice guys finish last? *International Journal of Organizational Analysis*, *11*, 137-152.
- Callinan, M., & Robertson, I.T. (2000). Work sample testing. *International Journal of Selection and Assessment*, *8*, 248-260.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2<sup>nd</sup> ed., Vol.1, pp. 687-731). Palo Alto, CA: Consulting Psychologist Press.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2<sup>nd</sup> ed., Vol. 1, pp. 687-731). Palo Alto, CA: Consulting Psychologists Press.
- Campbell, J. P., Gasser, M. B. & Oswald, R.L. (1996). The substantive nature of job performance variability. In K. Murphy (Ed.), *Individual difference and behavior in organizations* (pp. 258-259). San Francisco: Jossey Bass.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.). New York: Academic Press.
- Cook, M. (2004). *Personnel Selection: Adding value through people* (4<sup>th</sup> ed.). Chichester: Wiley.
- Goldstein, H.W., Yusko, K.P., Braverman, E.P., Smith, D.B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology*, *51*, 357-374.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Hillsdale, NJ: Erlbaum.
- Hattrup, K., O'Connell, M. S., & Wingate, P. H. (1998). Prediction of multidimensional criteria: Distinguishing task and contextual performance. *Human Performance*, *11*(4), 305-319.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004).

- Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639-683.
- Heneman, H.G., & Judge T. (2003). *Staffing organizations* (4th ed.). Irwin: McGraw-Hill.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment*, *9*, 152-194.
- Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257-266). Hillsdale, NJ: Erlbaum.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, *96*(1), 72-98.
- Kolz, A. R., McFarland, L. A., & Silverman, S. B. (1998). Cognitive ability and job experience as predictors of work performance. *Journal of Psychology*, *132*(5), 539-548.
- McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology*, *79*(4), 493-505.
- Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics and team work knowledge. *Personnel Psychology*, *58*, 583-611.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, *79*, 475-480.
- Murphy, K. R., & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: multivariate frameworks for studying test validity. *Personnel Psychology*, *50*(4), 823-854.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, *11*, 245-269.
- Ployhart, R. E., Schneider B, Schmitt N. (in press). *Staffing organizations: Contemporary practice and theory* (3<sup>rd</sup> ed.). Mahwah, NJ: Erlbaum.
- Quinones, M. A., Ford, J. K., & Teachout, M. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology*, *48*, 887-910.
- Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, *55*, 171-183.
- Robertson, T. R., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, *74*, 441-472.
- Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: updating and integrating some classic literature. *Journal of Personnel Psychology*, *58*(4), 1009-1037.
- Roth, P.L., Bobko, P., McFarland, L. A., & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of Black-White differences in overall and exercises scores. *Personnel Psychology*, *61*(3), 637-662.
- Russell, C.J. & Dean, M.A. (1994, August). *The effect of history on meta-analytic results: An example from personnel selection research*. Presented at the annual meetings of the Academy of Management, Dallas, TX.
- Salgado, J. F., Viswesvaran, C. & Ones, D.S. (2001). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D. S. Ones, H.K. Sinangil & C. Viswesvaran (Eds.), *International Handbook of Work and Organizational Psychology*. (Vol. 1, pp. 165-199). London, UK: Sage.
- Schmidt, F. L., & Hunter, J. E. (1992). Causal modeling processes determining job performance. *Current Directions in Psychological Science*, *1*, 89-92.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.
- Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). The impact of job experience and ability on job knowledge, work sample performance and supervisory ratings of job performance. *Journal of Applied Psychology*, *71*, 432-439.
- Schmitt, N., Clause, C.S., & Pulakos, E.D. (1996). Subgroup differences associated with different measures of some common job relevant constructs. In C.L. Cooper & I.T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology*, (pp. 115 - 140). New York: Wiley.
- Schmitt, N., Gooding, R. Z., Noe, R.A., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, *37*, 407-422.
- Schneider B, Schmitt N. (1986). *Staffing organizations* (2nd ed.). Glenview, IL: Scott, Foresman.
- Speier, C., & Frese, M. (1997). Generalized self efficacy as mediator and moderator between control and complexity at work and personal initiative: A longitudinal field study in East Germany. *Human Performance*, *10*(2), 171-192.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.
- Terpstra, D.E., Kethley, R.B., & Foley, R.T. (2000). The nature of litigation surrounding five screening devices. *Public Personnel Management*, *29*, 43-54.
- Tesluk, P. E., & Jacobs, R. J. (1998). Toward an integrated model of work experience. *Personnel Psychology*, *51*, 321-355.
- Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology*, *61*, 871-925.
- Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of con-

textual performance. *Journal of Applied Psychology*, 81(5), 525-531.

Viswesvaran, C. (2001). Assessment of individual job performance: A review of the past century and a look ahead.

In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 110-126). Thousand Oaks, CA: Sage.

Manuscrito Recibido 21/1/2009

Revisión Recibida 20/3/2009

Aceptado 26/3/2009